THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 55, No. 9, November 1976
Printed in U.S.A.

Speech Encryption by Manipulations of LPC Parameters

By M. R. SAMBUR and N. S. JAYANT

(Manuscript received May 14, 1976)

This paper discusses several manipulations of LPC (linear predictive coding) parameters for providing speech encryption. Specifically, the paper considers temporal rearrangement or scrambling of the LPC code sequence, as well as the alternative of perturbing individual samples in the sequence by means of pseudo-random additive or multiplicative noise. The latter approach is believed to have greater encryption potential than the temporal scrambling technique, in terms of the time needed to "break the secrecy code." The encryption techniques are assessed on the basis of perceptual experiments, as well as by means of a quantitative assessment of speech-spectrum distortion, as given by an appropriate "distance" measure.

I. INTRODUCTION

Encryption can be an important requirement in speech communication systems. Conventionally, encryption has largely been accomplished by signal manipulations in the frequency domain; for example, by means of spectrum inversion techniques.[1] With the increased popularity of digital codes for speech transmission, time-domain encryption techniques have received increased attention. Typically the time-domain encryption technique consists of temporal rearrangement of samples within a time block. For the scrambling of PCM bits in speech waveform coding, a block-length that is at least a pitch period long is usually adequate to provide a nonspeech-like output waveform. Similarly, the scrambling of differential PCM and delta-modulation bits can also produce a nonspeech-like output waveform provided that the time-block is sufficiently long. For example, in a 24-kb/s speech code, this constraint implies approximately a block length of 64 samples for an adequate scrambling of the coded bits.[2]

The temporal scrambling of speech samples within millisecond-length blocks generally provides what may be referred to as casual encryption. This means that a noncasual 'eavesdropper' can break the

speech secrecy code by the simple expedient of running tbrough a finite number of possible rearrangements of the disarranged speech samples that are received. Greater degrees of encryption or secrecy can be achieved by employing mucb longer speech blocks for scrambling or, alternatively, by subjecting individual speecb samples to pseudo-random additive or multiplicative perturbations whose undoing is typically more time-consuming than a simple temporal rearrangement of clean digits or bits.

The purpose of this paper is to point out tbat casual encryption as well as more formal secrecy can be achieved by appropriate manipulations of the linear predictive coding (LPC) parameters.[3,4] The use of an LPC code is by no means a necessary requirement for encryption; it can be achieved in conjunction with any kind of speech digitizers, such as the waveform codes[5] discussed above. However, when the channel capacity of communication systems dictates a low-bit-rate vocoder instead of a generally bigher-bit-rate waveform code, tbe LPC parameter manipulations discussed in this paper may provide a naturally appropriate basis for speech encryption and/or secrecy. It shall also be seen that an efficient encryption of the LPC parameters can be achieved more readily tban similar techniques used to encrypt waveform codes. For example, an adequate block length for scrambling the LPC parameters can be as sbort as 6 to 8 samples, while tbe block length for waveform scrambling is typically 16 to 64 samples.

In this paper, Section II provides a brief description of LPC encoding of speech, while Section III considers the use of temporal scrambling and pseudo-random sample perturbations for casual and formal encryption in the LPC domain. Section IV describes attempts to measure the efficacy of tbe encryption tecbniques. These measurements involved informal perceptual experiments (tbe results are usually unambiguous and one-dimensional enough not to require formal subjective testing), as well as a comparison of alternative techniques in terms of speech-spectrum distortions that they provide. The spectrum distortion was assessed by an appropriate distance measurement. This distance approach has the advantage of being quantitative; however, as discussed in Section IV, the distance criterion has to be invoked with caution because spectral distortion, as such, is not a definitive measure of speech encryption.

II. LINEAR PREDICTION SPEECH MODEL

Tbe metbod of linear prediction has proved quite popular and successful for use in speech-compression systems.[3,6,7] In this method, speech is modeled as the output of an all-pole filter $H(z)$ tbat is excited by a sequence of pulses separated by the pitch period for voiced sounds

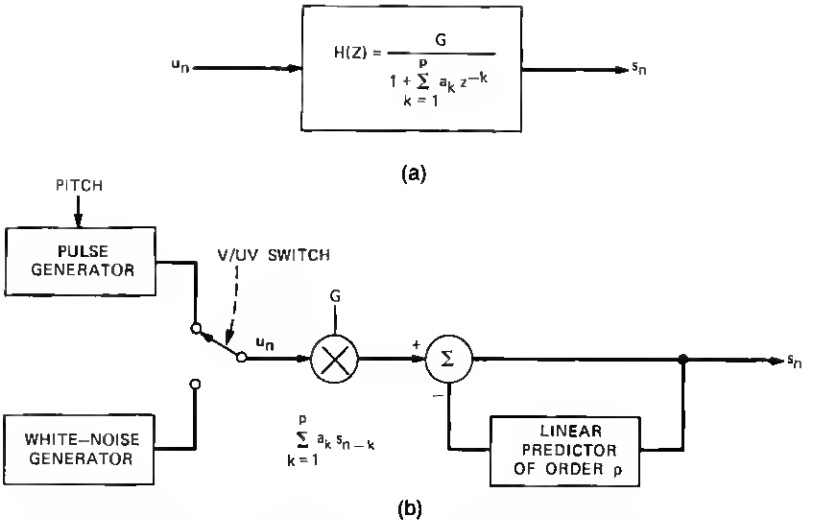


Fig. 1—Discrete model of speech production as employed in linear prediction. (a) Frequency-domain model. (b) Time-domain model.

or pseudo-random noise for unvoiced sounds. These assumptions imply that within a frame of speech, the output speech sequence is given by

$$s_n = \sum_{k=1}^{p} a_k s_{n-k} + G u_n,$$

where p is the number of modeled poles, u_n is the appropriate input excitation, G is the gain of the filter, and the a_k's are the coefficients characterizing the filter (linear prediction coefficients). Figure 1 illustrates the frequency-domain as well as the equivalent time-domain model of linear prediction speech production. To account for the nonstationary character of the speech waveform, the parameters a_k of the modeled filter are periodically updated during successive speech frames.* Generation of speech in this method requires a knowledge of the pitch, the filter parameters, and the gain of the filter (amplitude of excitation) in each speech frame.

The LPC coefficients model the combined effects of the vocal tract, glottal source, and radiation load in each frame of speech. Manipulations of the LPC coefficients can seriously perturb the frequency character of the speech signal and, hence, destroy the linguistic information present in the signal. In contrast, the measurements of pitch and gain represent the prosodic aspects of the speech and some characteristics

* A frame is a segment of speech thought adequate to assume stationarity of the speech process. Typical frame lengths employed range from 10 to 30 ms.

of the speaker. Manipulations of pitch and gain parameters will affect the prosody of the speech, hut not seriously diminish the linguistic aspects of the waveform. In Section III, we consider several methods for efficiently manipulating the LPC coefficients so as to encrypt the speech signal.

Since the purpose of this paper is the consideration of encryption techniques for low-hit-rate vocoders (2.4 kh/s or less), the manipulation schemes discussed in Section III were not performed directly on the LPC coefficients, hut rather on more desirahle alternate representations of these coefficients. The stability of the linear-prediction filter, $H(z)$, is extremely sensitive to small perturhations in the LPC coefficients and, thus, it is not possihle to achieve low-hit-rate coding by transmitting the LPC coefficients.[6] However, by transmitting either the log area coefficients or the parcor coefficients, a 2.4-kb/s vocoder is readily achieved.[6] The log area coefficients are nonlinearly related to the LPC coefficients by

$$g_i = \log \frac{1 + k_i}{1 - k_i},$$

where the k_i's are termed the parcor coefficients.[7] If we denote $a_i^{(j)}$ as the ith linear prediction coefficient for a jth-pole linear-prediction model, then

$$k_i = a_i^{(i)}.$$

The parcor coefficients have the very important property that if

$$|k_i| < 1, \qquad i = 1, \cdots, p,$$

then it is guaranteed that the linear prediction filter is stahle.[4] Thus, small perturhations in the parcor coefficients or log-area coefficients will not affect the stahility of the modeled filter.

III. ENCRYPTION TECHNIQUES

3.1 Temporal scrambling

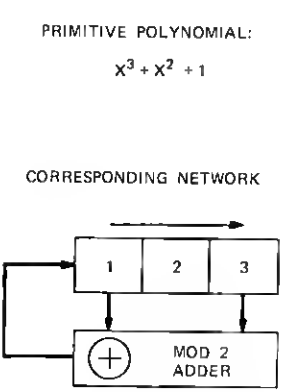
The rearrangement of samples within a hlock of length L is achieved hy assigning to each sample a new address A ($A = 1$, or 2, or 3, $\cdots$, or L) as determined by the state of a maximal-length shift-register arrangement. The theory and design of maximal length sequences is well documented.[8,9] Here, we simply provide a constructive recapitulation for the purpose of this paper. The idea is to start with a shift register whose length is $D = \log_2 L$ (assume that the hlock length is a power of 2, and that elements in the register are either 1 or 0). The next step is to select a so-called primitive polynomial $P_D(x)$ of degree D, and to include stage $(D - S)$ in the register ($S = 0$ to $D - 1$) in

an exclusive OR (modulo 2 add) feedback arrangement, if the coefficient of x^S in $P(x)$ is nonzero. The resulting network now generates a succession of $2^D - 1 = L - 1$ nonzero states in the shift register at successive 'clock' times, after which the cycle repeats, starting once again with the original initial state of the shift register. The number of nonzero states in the cycle is identically equal to the repetition period $L - 1$ of the cycle. Consequently, the $L - 1$ states of the shift-register (specifically, their decimal equivalents) can be utilized as "pseudo-random" addresses for a block of $L - 1$ input samples in a one-to-one mapping of addresses. If the input block has L rather than $L - 1$ samples (because of the frequent requirement that L be a power of 2), the address of the Lth sample is usually left unaltered by the scrambler. Such simplicity is not, however, inevitable, and appropriate manipulations that scramble all L samples are quite conceivable.

Figure 2 illustrates the scrambler design for the example of $D = 3$ and $L = 7$, as defined by a primitive polynomial $P_3(X) = X^3 + X^2 + 1$. It is seen how input samples (1, 2, 3, 4, 5, 6, and 7) get scrambled into the pseudo-random positions (1, 4, 6, 7, 3, 5, and 2) in a reversible one-to-one mapping.

Figure 3 illustrates an alternative design, as defined by a second primitive polynomial of degree 3, $P_3(X) = X^3 + X + 1$. In this case, the output addresses of the input samples are the positions (1, 4, 2, 5, 6, 7, and 3).

It is clear that in each of the arrangements in Figs. 2 and 3, the use of a different initializing sequence (other than 001) can lead to a totally different mapping of sample addresses. There would be $L - 1$ nonzero initializations, corresponding to every given $P_3(X)$. Incidentally, the number of primitive polynomials of degree 3 is 2.



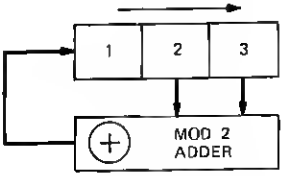
NUMBER	STATE OF SHIFT REGISTER	DECIMAL EQUIVALENT
1	001	1
2	100	4
3	110	6
4	111	7
5	011	3
6	101	5
7	010	2
↑	001	↑
INPUT ADDRESSES	REPETITION PERIOD = 7	OUTPUT ADDRESSES

Fig. 2—Scrambler design with a three-stage shift register.

PRIMITIVE POLYNOMIAL:

$X^3 + X + 1$

CORRESPONDING NETWORK



NUMBER	STATE OF SHIFT REGISTER	DECIMAL EQUIVALENT
1	001	1
2	100	4
3	010	2
4	101	5
5	110	6
6	111	7
7	011	3
↑	001	↑
INPUT ADDRESSES	REPETITION PERIOD = 7	OUTPUT ADDRESSES

Fig. 3—Alternate scrambler design with a three-stage shift register.

Table I lists for $D = 1$ to 12 a typical set of primitive polynomials and also the number of primitive polynomials for each D. Note, for example, that a 12-stage shift register with an exclusive OR feedback network involving stages 12, 11, 8, and 6 provides one of 144 possible bases for a scrambler that would operate on an input block of $2^{12} = 4096$ samples.

The possibility of alternate scrambler designs (as defined by different initializations and/or different primitive polynomials) is an important consideration from the point of view of the average descrambling time needed for an eavesdropping code-breaker.

3.2 LPC parameter scrambling

The effectiveness of any scrambling scheme in perturbing the sequence of samples is directly proportional to the lack of similarity or

Table I—List of primitive polynomials

Degree D	Typical Primitive Polynomial	Number of Primitive Polynomials of Degree D
1	$X + 1$	1
2	$X^2 + X + 1$	1
3	$X^3 + X + 1$	2
4	$X^4 + X + 1$	2
5	$X^5 + X^2 + 1$	6
6	$X^6 + X + 1$	6
7	$X^7 + X + 1$	18
8	$X^8 + X^4 + X^3 + X^2 + 1$	16
9	$X^9 + X^4 + 1$	48
10	$X^{10} + X^3 + 1$	60
11	$X^{11} + X^2 + 1$	176
12	$X^{12} + X^6 + X^4 + X + 1$	144

dynamic ranges of the samples to be scrambled. The greater the range of values assumed by the samples, the more effeetive the scramhling scheme.[1] For an efficient scrambling of the LPC parameters, let us begin by ordering the parameters in the following manner: the first sample in the first block is x_{11}, where x_{in} denotes the ith LPC parameter* in the nth analysis frame. The second sample is x_{21} and the third sample is x_{31}. The arrangement proceeds in this fashion until the $(p + 1)$th† sample, which is defined as x_{12}. Thus, the ordering of LPC parameters for purposes of scrambling is simply a concatenation of the p LPC parameters in eacb sequential analysis frame.

Using this particular arrangement, it can be seen that within a block of data there is a wide distribution of values assumed by tbe various samples. Tbis observation follows from the fact that the measured LPC parameters for any given analysis frame will usually vary across the entire permissible range of values. For example, the p measured values of the parcor coefficients in any given frame will typically he somewhat uniformly spread across the permissihle range of -1 to 1.[4] The particular arrangement of LPC parameters given ahove will tbus be effective for scramhling purposes due to the large resulting dynamic range. In Section IV, we show that a block length as small as eigbt samples ($L = 8$) is sufficient to destroy the linguistie information in the synthetic signal produced by a 12th order analysis ($p = 12$).

3.3 Pseudo-random perturbations

For a more secure secrecy coding of the speech signal, the LPC parameters can be modified by a pseudo-random additive or multiplicative perturhation. Since the repetitive period of any typical pseudo-random numher generator is extremely large, the proeess of undoing or breaking the encryption is quite difficult and time-consuming.

Since one of the goals of the present study was to perceptually assess the linguistic information in the synthesized speech generated by the encrypted LPC parameters, the pseudo-random number perturhation scheme was designed to retain the stahility of the modified LPC filter. Thus, for tbe manipulation of the parcor coefficients, the pseudo-random numher technique involved the transmission of the sequence of parameters

$$y_{in} = k_{in} \times r_{in},$$

where

k_{in} = ith pareor coefficient in nth frame

r_{in} = ith pseudo-random number in nth frame; $|r_{in}| \leqq 1$.

* The LPC parameters considered in this paper are either the log-area coefficients or parcor coefficients.

† p = order of LPC analysis.

Since $|r_{in}| \leq 1$, $|y_{in}| < 1$ and tbe stability of tbe LPC filter is guaranteed. For the modification of tbe log-area coefficients, tbe technique is simply to transmit

$$y_{in} = g_{in} + r_{in}.$$

The stahility of the resulting LPC filter is guaranteed regardless of tbe range of r_{in}. This result follows from the fact that any real value of y_{in} will lead to parcor parameters that are less tban 1.

In viewing the pscudo-random number manipulation of the LPC parameters, it should be noted that the spectral cbaracteristics of tbe LPC filter are more sensitive to cbanges in the parcor coefficients tban to changes in the log-area coefficients.[10] Thus, manipulation of the parcor coefficients is a more direct and efficient technique for perturhing the spectral properties of the LPC filter. For tbis reason the pseudo-random techniques discussed in this paper were applied only to the parcor coefficients. If pseudo-random number manipulation is to be applied to the log-area coefficients, the manipulation can be made most effective if tbe probability distribution of the random number generator is nonuniform, in order to mimic tbat of tbe log-area coefficient.[10]

For tbe experimental examination of the pseudo-random number perturbation of the parcor coefficients, tbe following two probability distrihutions were used for generating r_{in}:

(*i*) r_{in} was uniformly distrihuted between -1 and 1, or

(*ii*) r_{in} was, with equal probability, set to -1 or 1.

The second distribution was studied hecause the resulting manipulation of the parcor coefficients is particularly easy to perform and, as we shall soon discuss, is effective in destroying the intelligibility of the encrypted speech. However, tbe "hreaking" of the encryption coding using the second distrihution is not difficult to achieve by using the availahle knowledge of the statistical range of the parameters. For example, it is well known that the first parcor coefficient is almost always positive.[4] Thus, a negative value of the first parcor coefficient indicates a manipulation of this parameter. If the listener knows that a $+1$ or -1 manipulation of the parameters is heing employed, tben a simple reversal of sign hreaks the encryption.

IV. EXPERIMENTAL STUDY

In this section, we examine the effectiveness of tbe various encryption tecbniques in destroying tbe intelligibility of the output speech signal. For this purpose, an informal perceptual evaluation was conducted. To evaluate objectively the efficacy of the tecbniques, an LPC distance measure proposed by Itakura[11] was used to reinforce and supplement

the perceptual examination. Before discussing the LPC distance measure, we emphasize that *this measure may not be a definitive or complete description of encryption efficiency;* but it is a good measure of spectral distortion, which in turn turns out to be a useful (if not ideal) indicator of intelligibility loss.

4.1 Distance measure

The LPC distance measure is defined as

$$d_n = \ln (\mathbf{a}_n V \mathbf{a}_n^T / \mathbf{b}_n V \mathbf{b}_n^T),$$

where

$\mathbf{a}_n$ = Original LPC coefficient vector $(1, a_1, \cdots, a_p)$ measured in the nth frame of the speech signal.

$\mathbf{b}_n$ = LPC coefficient vector determined after manipulation of the original parameters in the nth frame

and

$$V = [v(|i - j|)], \qquad (i, j = 0, 1, \cdots, p),$$

where $v(i)$ are the normalized correlation coefficients that are computed directly from $\mathbf{b}_n$.[3,10]

The measure d_n has been very effectively applied in problems of speech recognition,[11] speaker recognition,[12] and variable frame-rate synthesis.[13,14] Gray and Markel[15] have recently demonstrated that the measure d_n is very closely related to the rms spectral distance measure. Sambur and Jayant[16] have also studied the significance of the measure, and a complete discussion of the utility of the measure for assessing spectral distortions can be found in their paper. For purposes of this paper, the important facts to appreciate about the measure d_n are

(*i*) The greater the value of d_n, the more pronounced the spectral distortions of the original sound.

(*ii*) A value of $d_n = 0.9$ is a "perceptually" significant boundary for evaluating spectral distortion.[13]

4.2 Experiment

For the experimental study, four sentences spoken by four different speakers were analyzed using a 12th order ($p = 12$) LPC autocorrelation analysis for each contiguous 20-ms frame. The sentences analyzed were:

(*i*) A lathe is a big tool.

(*ii*) May we all learn a yellow lion roar.

(*iii*) Few thieves are never sent to the jug.

(*iv*) It's time we rounded up that herd of Asian cattle.

The encryption schemes that were formally evaluated both perceptually and with the distance measure of Section 4.1 were:

a. Scrambling
   (1) Block length = 16
   (2) Block length = 8
b. Pseudo-random manipulation*
   (1) Uniform distribution of r_{in} for $i = 1$ and $r_{in} = 1$ for $i > 1$.
   (2) Uniform distribution of r_{in} for $i \leqq 6$ and $r_{in} = 1$ for $i > 6$.
   (3) Uniform distribution of r_{in} for all i $(1 \leqq i \leqq 12)$.
   (4) ± 1 distribution of r_{in} for $i = 1$ and $r_{in} = 1$ for $i > 1$.
   (5) ± 1 distribution of r_{in} for $i \leqq 6$ and $r_{in} = 1$ for $i > 6$.
   (6) ± 1 distribution of r_{in} for all i $(1 \leqq i \leqq 12)$.

Experiment b was performed to determine the number of parcor coefficients that must be altered to effectively encrypt the signal. Since the parcor coefficients are approximately ordered in terms of their spectral sensitivity,[4] these experiments were performed by sequentially removing from manipulation the less sensitive parameters.

4.3 Results

4.3.1 Distance evaluation

4.3.1.1 Uniform pseudo-random manipulation. Figure 4 illustrates the distance-evaluation of the sentence "May we all learn a yellow lion roar" for the uniform pseudo-random number manipulation of the parcor coefficients. Parts (a), (b), and (c) of Fig. 4 indicate, respectively, the results of experiments b(1), b(2), and b(3) of Section 4.2. The straight solid line in each part of the figure depicts the perceptually significant threshold for assessing spectral distortions ($d = 0.9$). Any frame with a distance larger than the threshold is perceptually different from the nonencrypted speech. To show just how dramatically the perturbation in the spectral character of the speech can be, Fig. 5 illustrates the calculated linear prediction spectrum (dotted line) for the nonencrypted speech frame and the corresponding linear prediction spectrum (solid line) for the same frame of encrypted speech. The measured LPC distance between the illustrated spectra is $d_n = 3.0$, or approximately the average value of distance for uniform pseudo-random manipulation of the first coefficient. From this figure, it can be expected that the character of the encrypted speech is completely different from that of the original speech.

* Remember r_{in} denotes the pseudo-random number multiplicative factor for the ith parcor coefficient in the nth frame.

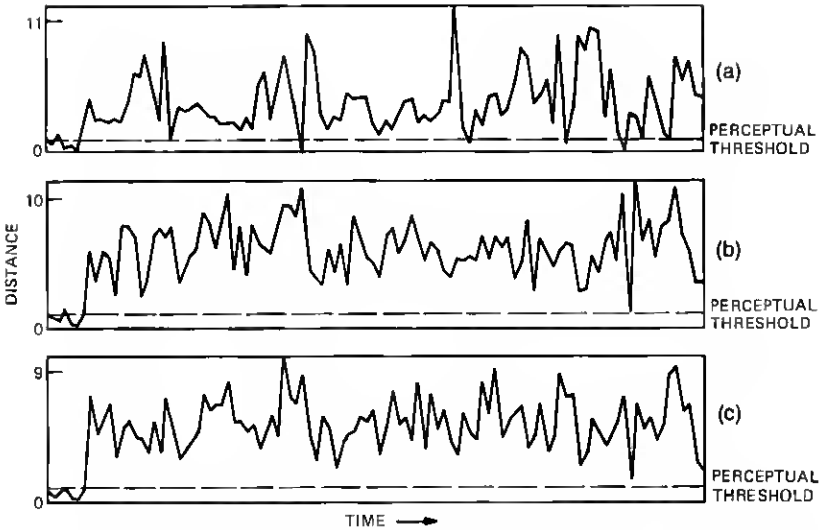


Fig. 4—LPC distance as a function of time across the utterance, "May we all learn a yellow lion roar," for uniform pseudo-random perturbation of the parcor parameters. (a) Manipulation of k_1; average distance = 3.4. (b) Manipulation of k_1 to k_6; average distance = 4.4. (c) Manipulation of all k_i; average distance = 4.4.

The results depicted in Fig. 4 are typical of the distance evaluation results for the uniform pseudo-random manipulation of the parcor coefficients determined for the other sentences examined. It is interesting to note that the average distance for an encryption scheme that manipulates the first six parameters is not significantly lower than the average distance obtained for the manipulation of all 12 parameters. This result can be anticipated from the fact that the higher-ordered parcor coefficients are much less sensitive than the lower-ordered parameters, and changes in these higher-ordered parameters do not significantly change the spectral character of the sound.[4] Thus, a less-expensive and equally effective encryption scheme can be obtained by manipulating only a few lower-ordered parameters. To determine the optimum number of parameters necessary for an efficient, uniform,

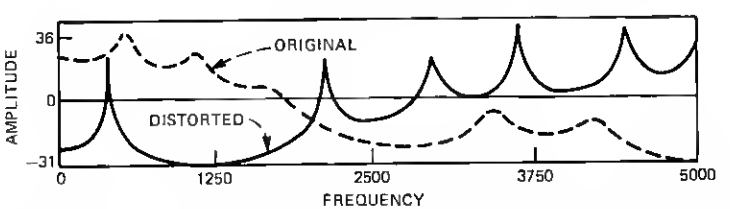


Fig. 5—Comparison of the distorted LPC spectra and the original LPC spectrum. Distance between the spectrum equals 3.0.

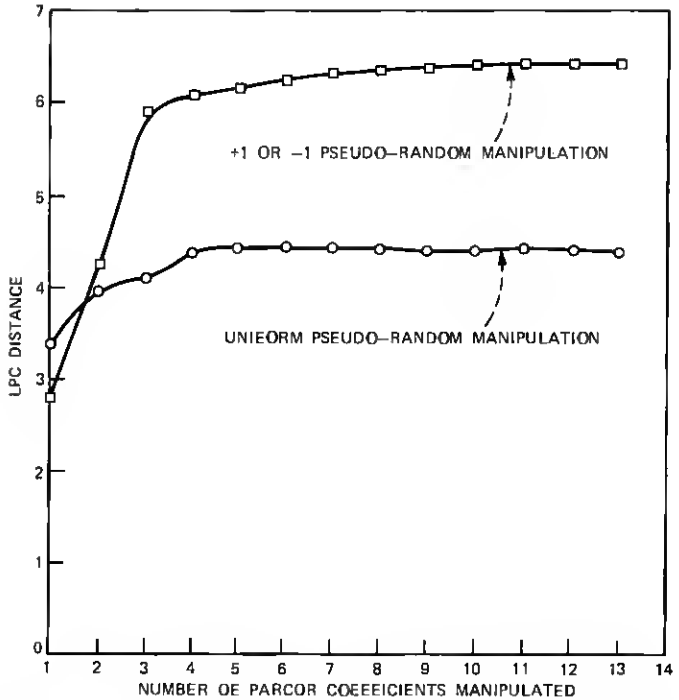


Fig. 6—Average LPC distance as a function of number of parcor coefficients manipulated by pseudo-random number techniques.

pseudo-random encryption, we sequentially increased the number of parcor parameters perturbed by uniform pseudo-random manipulation and measured the average LPC distance. Figure 6 illustrates the average LPC distance as a function of the number of parameters manipulated. From this figure, it can be seen that a scheme that perturbs only the first four parcor coefficients is quite efficient.

4.3.1.2 Pseudo-random manipulation of +1 or −1. Figure 7 shows the detailed distance-evaluation scores for the $+1$ or -1 pseudo-random perturbation of the sentence "May we all learn a yellow lion roar." Parts (a), (b), and (c) of the figure correspond to experiments b(4), b(5), and b(6), respectively. Figure 6 illustrates the average LPC distances obtained for encryption schemes that sequentially increase the number of parameters subjected to $+1$ or -1 pseudo-random manipulations. We note from Figs. 6 and 7 that again the perturbation of the higher-ordered parcor coefficients does not significantly add to the effectiveness of the encryption scheme. It can also be seen from these figures that $+1$ or -1 pseudo-random manipulation is generally superior (except for the manipulation of only k_1) to the uniform pseudo-random number scheme in distorting the speech signal. How-

ever, as noted previously, this form of encryption is easier to hreak than uniform pseudo-random numher coding.

4.3.1.3 Scrambling. Figure 8 shows the frame-hy-frame distance scores for the scrambling of the parcor coefficients for the sentence "May we all learn a yellow lion roar." The illustrated results are typical of the results obtained for the other analyzed sentences. A comparison of the distances results of the pseudo-random schemes (Fig. 6) shows that a scrambling encryption with a hlock length of only eight samples ($L = 8$) is at least as effective in distorting the spectral properties of the original signal as a pseudo-random manipulation of the first parcor coefficient. A scramhling scheme with a hlock length of 16 ($L = 16$) or more samples is superior to any of the pseudo-random schemes studied. It is interesting to note that the scrambling manipulation saturates in effectiveness for hlock length greater than 16. Since the range of the parcor coefficients is confined to $-1 \leqq k_i \leqq 1$, increasing the hlock length heyond 16 does not increase the dynamic range of the sample within the hlock and, thus, the effectiveness of the scramhling is not cnhanced for $L > p$ (see Section 3.2).

4.3.2 Perceptual evaluation

To support the results of the distance study, the various encrypted utterances were presented to a group of listeners for an informal perceptual evaluation of the manipulation schemes. To avoid any

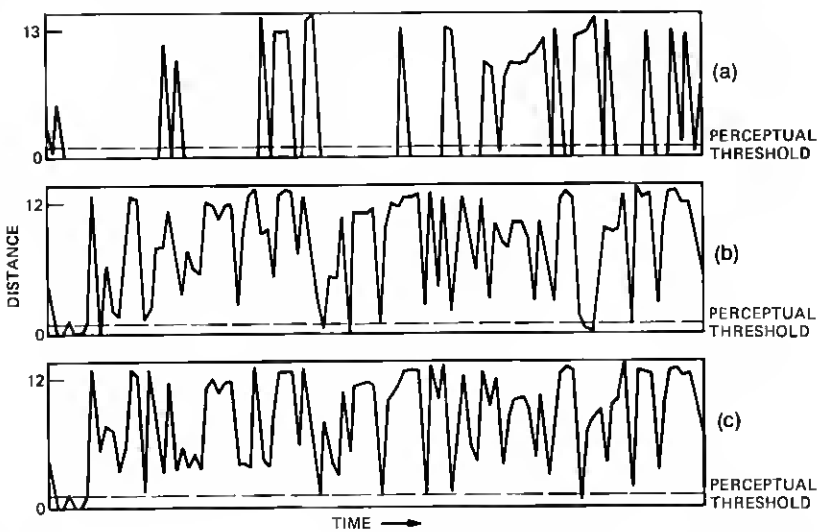


Fig. 7—LPC distance as a function of time across the utterance "May we all learn a yellow lion roar" for the +1 or −1 pseudo-random manipulation of the parcor coefficient. (a) Manipulation of k_1; average distance = 2.8. (b) Manipulation of k_1 to k_6; average distance = 6.2. (c) Manipulation of all k_i; average distance = 6.3.

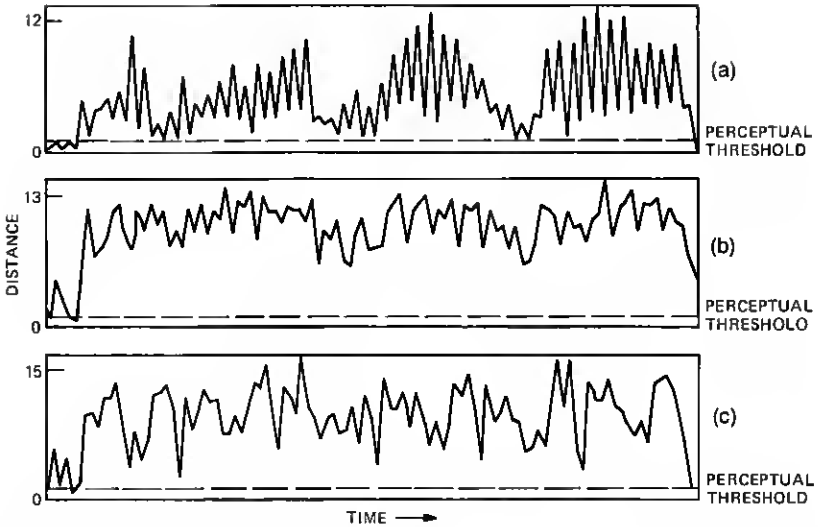


Fig. 8—LPC distance as a function of time across the utterance "May we all learn a yellow lion roar" for the scrambling of the parcor coefficient. (a) Block length = 8; average distance = 3.8. (b) Block length = 16; average distance = 7.7; (c) Block length = 64; average distance = 7.6.

prohlems posed hy the awkward linguistic content of the analyzed sentences, the listeners in this study were all familiar with the sentences, and were also informed that the utterances to he heard were typical sentences used to evaluate vocoder systems.

The listeners in the experiment were asked to determine the intelligihility of the utterance and to rank-order the effectiveness of the encryption schemes. For all the techniques studied, except for the $+1$ or -1 manipulation of only k_1, the listeners unanimously agreed that the encrypted utterances were clearly nonspeech-like. However, for the uniform pseudo-random techniques manipulating only the first parcor coefficients, the listeners noted that, even though the complete utterances could not he understood, there were certain instances in the encrypted utterances that were somewhat speech-like and understandahle. These instances prohahly correspond to points in the encrypted speech for which the LPC distances fall helow the perceptual threshold. In characterizing the nonspeech-like quality of the encrypted utterances, the listeners termed the pseudo-random perturhed utterances as sounding like "one continuous huzz;" the scramhled utterances sounded like "water running through a pipe."

In rank-ordering the encryption schemes, the listeners were quite definite in characterizing the $+1$ or -1 psuedo-random manipulation of only the first parcor coefficient as least effective. The scramhling

with block length of 16 ($\bar{d} = 7.7$) was ranked about equal to the $+1$ or -1 pseudo-random manipulation of all 12 parcor coefficients ($\bar{d} = 6.3$), and also to the same manipulation of only the first six coefficients ($\bar{d} = 6.2$). The uniform pseudo-random scheme that altered all 12 coefficients ($\bar{d} = 4.4$) was ranked equal to the scheme that perturbed only the first six coefficients ($\bar{d} = 4.4$), and both techniques were ranked slightly less effective than the scrambling with block length of 16 ($\bar{d} = 7.7$) and the equivalent $+1$ or -1 pseudo-random schemes. The other techniques were ranked somewhere in the middle. The perceptual rank-ordering of the various manipulation schemes corresponded almost exactly to the distance evaluation and, thus, reinforced the conclusions in that evaluation.

V. CONCLUSIONS

There is great interest in low-bit-rate speech-transmission systems and in the "securing" of these transmission systems. The purpose of this paper is to investigate various methods for encrypting a low-bit-rate LPC transmission system. The methods chosen for investigation were schemes that either scrambled the string of input parcor coefficients or multiplied the coefficients by a pseudo-random number. The schemes were evaluated by an informal perceptual experiment and by the use of an LPC distance measure. The results of the evaluations suggest that all the schemes are somewhat successful in distorting the original signal. The most successful scheme was the scrambling technique with a block length of 16 samples. The pseudo-random manipulations were almost as effective.

In viewing the results of the evaluations, it is important to note that the distortion of the speech signal is only one consideration in designing an encryption system. Another consideration is the difficulty of "breaking" the security code. Of the codes examined, the uniform pseudo-random number manipulation is the most difficult to break. The scrambling scheme is the next most difficult and the $+1$ or -1 pseudo-random scheme is the easiest. Still another consideration is the transmitter-end complexity of the encryption scheme. Although this complexity is somewhat difficult to assess, it appears that the scrambling scheme is the least complex and the uniform pseudo-random manipulation is the most complex. In choosing any of these encryption schemes, a user would balance the various merits and liabilities of the techniques.

VI. ACKNOWLEDGMENT

The manipulation of LPC parameters was suggested by Professor B. S. Ramakrishna of the Electrical Communication Engineering

Department, Indian Institute of Science, Bangalore, during a visit by one of the authors (N. S. Jayant).

REFERENCES

1. D. Kahn, *The Code Breakers*, New York: Macmillan, 1967, pp. 551–554.
2. N. S. Jayant, "Step-size Transmitting Differential Coders for Mobile Telephony," B.S.T.J., *54*, No. 9 (Novemher 1975), pp. 1557–1581.
3. B. S. Atal and S. L. Hanauer, "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave," J. Acoust. Soc. Amer., *50* (1971), pp. 637–655.
4. J. D. Markel, A. H. Gray, Jr., and H. Wakita, "Linear Prediction of Speech-Theory and Practice," Speech Communication Research Laboratory, Inc., Santa Barhara, Calif., Monograph 10, 1973.
5. N. S. Jayant, "Digital Coding of Speech Waveforms—PCM, DPCM, and DM Quantizers," Proc. IEEE (May 1974), pp. 611–632.
6. M. R. Samhur, "An Efficient Linear Prediction Vocoder," B.S.T.J., *54*, No. 10 (Decemher 1975), pp. 1693–1723.
7. F. Itakura et al., "An Audio Response Unit Based on Partial Autocorrelation," IEEE Trans. Commun., *COM-20*, No. 4 (August 1972), pp. 792–797.
8. S. Golomh, *Shift Register Sequences*, San Francisco: Holden Day, 1967.
9. R. G. Gallager, *Information Theory and Reliable Communications*, New York: John Wiley, 1968.
10. J. Makhoul and R. Viswanathan, "Quantization Properties of Transmission Parameters in Linear Predictive Systems," Bolt Beranek and Newman, Inc., Report No. 2800, April 1974.
11. F. Itakura, "Minimum Prediction Residual Principle Applied to Speech Recognition," IEEE Trans. Acoust., Speech and Signal Proc., *ASSP-23*, No. 1 (Fehruary 1975), pp. 67–72.
12 H. Wakita, "On the Use of Linear Prediction Error Energy for Speech and Speaker Recognition," J.A.S.A., *57*, Supplement No. 1 (Spring 1975). (A)
13. D. T. Magill, "Adaptive Speech Compression for Packet Communication Systems," Telecommunications Conference Record, IEEE Puhl. 73, CH0805-2, 29D 1–5.
14. J. R. Makhoul, L. Viswanathan, L. Cosel, and W. Russel, "Natural Communication with Computers: Speech Compression Research at BBN," BBN Report No. 2976, Vol. II, Bolt Beranek and Newman, Inc., Cambridge, Massachusetts, Decemher 1974.
15. A. H. Gray and J. D. Markel, "COSH Measure for Speech Processing," J.A.S.A., *58*, Supplement No. 1 (Fall 1975). (A)
16. M. R. Sambur and N. S. Jayant, "LPC Analysis/Synthesis From Speech Inputs Containing Noise or Additive White Noise," IEEE Trans. Acoust., Speech, and Signal Proc., *ASSP-24*, No. 6 (Decemher 1976).